

What is claimed is:

1 1. A system for creating query refinement suggestions, comprising:
2 a matcher matching at least one search document retrieved responsive to a
3 query to one or more stored queries; and
4 a scorer scoring the stored query as a potential query refinement
5 suggestion.

1 2. A system according to Claim 1, further comprising:
2 a document matcher matching the at least one search document to one or
3 more stored documents associated with the one or more stored queries.

1 3. A system according to Claim 1, further comprising:
2 a weight associated with at least one such stored query; and
3 a clusterer forming at least one cluster based on the stored query and
4 weight.

1 4. A system according to Claim 3, further comprising:
2 a term vector used in cluster formation computed from terms extracted
3 from the stored query and based on the weight for the stored query.

1 5. A system according to Claim 4, further comprising:
2 a distance for the term vector; and
3 the cluster forming the at least one cluster relative to the distance.

1 6. A system according to Claim 3, further comprising:
2 a ranker ranking the at least one cluster relative to the at least one other
3 cluster by evaluating a relevance score associated with each search document
4 corresponding to each matched stored document.

1 7. A system according to Claim 6, further comprising:
2 a selector selecting at least one ranked cluster as a potential refinement
3 cluster.

1 8. A system according to Claim 3, further comprising:

2 a centroid computed as a weighted center of the at least one cluster; and
3 a score computed for each stored query relative to the centroid.

1 9. A system according to Claim 8, further comprising:
2 a selector naming the at least one cluster for at least one scored stored
3 query.

1 10. A system according to Claim 1, further comprising:
2 a threshold applied to the stored scored query.

1 11. A system according to Claim 1, further comprising:
2 a precomputation engine associating one or more stored documents to the
3 stored query based on at least one of a chosen search document, a set of search
4 documents, regenerated previous search documents, and inverted cached
5 document and query pairings.

1 12. A method for creating query refinement suggestions, comprising:
2 matching at least one search document retrieved responsive to a query to
3 one or more stored queries; and
4 scoring the stored query as a potential query refinement suggestion.

1 13. A method according to Claim 12, further comprising:
2 matching the at least one search document to one or more stored
3 documents associated with the one or more stored queries.

1 14. A method according to Claim 12, further comprising:
2 associating a weight with at least one such stored query; and
3 forming at least one cluster based on the stored query and weight.

1 15. A method according to Claim 14, further comprising:
2 computing a term vector used in cluster formation from terms extracted
3 from the stored query and based on the weight for the stored query.

1 16. A method according to Claim 15, further comprising:
2 determining a distance for the term vector; and

- 3 forming the at least one cluster relative to the distance.
- 1 17. A method according to Claim 14, further comprising:
2 ranking the at least one cluster relative to the at least one other cluster by
3 evaluating a relevance score associated with each search document corresponding
4 to each matched stored document.
- 1 18. A method according to Claim 17, further comprising:
2 selecting at least one ranked cluster as a potential refinement cluster.
- 1 19. A method according to Claim 14, further comprising:
2 computing a centroid as a weighted center of the at least one cluster; and
3 computing a score for each stored query relative to the centroid.
- 1 20. A method according to Claim 19, further comprising:
2 naming the at least one cluster for at least one scored stored query.
- 1 21. A method according to Claim 12, further comprising:
2 applying a threshold to the stored scored query.
- 1 22. A method according to Claim 12, further comprising:
2 associating one or more stored documents to the stored query based on at
3 least one of a chosen search document, a set of search documents, regenerated
4 previous search documents, and inverted cached document and query pairings.
- 1 23. A computer-readable storage medium holding code for performing
2 the method according to Claim 12.
- 1 24. An apparatus for creating query refinement suggestions,
2 comprising:
3 means for matching at least one search document retrieved responsive to a
4 query to one or more stored queries; and
5 means for scoring the stored query as a potential query refinement
6 suggestion.

1 25. A system for providing search query refinements, comprising:
2 an associator associating a stored query and a stored document as a logical
3 pairing and assigning a weight to the logical pairing;
4 a searcher issuing the search query and producing a set of search
5 documents;
6 a matcher matching at least one search document to at least one stored
7 document and retrieving the stored query and the assigned weight associated with
8 the matching at least one stored document;
9 a clusterer forming at least one cluster based on the stored query and the
10 assigned weight associated with the matching at least one stored document; and
11 a scorer scoring the stored query associated with the matching at least one
12 stored document for the at least one cluster relative to at least one other cluster
13 and suggesting at least one such scored search query as a set of query refinements.

1 26. A system according to Claim 25, further comprising:
2 a selector selecting one such search document chosen from among the set
3 of search documents responsive to the search query issuance as the at least one
4 such search document.

1 27. A system according to Claim 25, further comprising:
2 a selector selecting the set of search documents as the at least one such
3 search document.

1 28. A system according to Claim 25, further comprising:
2 a query log tracking previous search queries; and
3 a regenerator regenerating a set of previous search documents produced by
4 the previous search queries as the at least one such search document.

1 29. A system according to Claim 25, further comprising:
2 a cache associating at least one cached document and one or more cached
3 queries as a cached pairing; and

4 an inverter inverting each cached pairing to associate at least one cached
5 query and one or more cached documents as the at least one such search
6 document.

1 30. A system according to Claim 25, wherein relevancy to the stored
2 query is estimated for the stored document as the weight assigned to the pairing.

1 31. A system according to Claim 30, wherein each such assigned
2 weight for a plurality of pairings corresponding to the stored query and the stored
3 document is summed.

1 32. A system according to Claim 25, wherein each stored query
2 comprises one or more terms, further comprising:
3 a term vector comprising the terms in the stored query associated with the
4 matching at least one stored document;
5 a distance determined for the term vector; and
6 the clusterer forming the at least one cluster relative to the distance.

1 33. A system according to Claim 32, further comprising:
2 a normalizer normalizing the term vector.

1 34. A system according to Claim 32, further comprising:
2 an evaluator computing a length of the term vector in multi-dimensional
3 space with each dimension equaling a sum of the weights of the term in a set of
4 associated stored queries.

1 35. A system according to Claim 32, further comprising:
2 a relevance score assigned to the at least one search document; and
3 a ranker ranking the at least one cluster relative to the at least one other
4 cluster by the relevance score associated with the matching at least one search
5 document and a number of the matching at least one search document.

1 36. A system according to Claim 35, further comprising:

2 a selector selecting one of more of the ranked at least one cluster as
3 potential refinement clusters based on the rankings.

1 37. A system according to Claim 36, further comprising:
2 a centroid computed as a weighted center for each such potential
3 refinement cluster; and
4 the scorer scoring the stored query associated with the matching at least
5 one stored document for the potential refinement cluster relative to the centroid.

1 38. A system according to Claim 37, further comprising:
2 an evaluator computing the centroid as a normalized sum of a product of
3 the term vector for each stored query and the relevance score associated with the
4 matching at least one search document.

1 39. A system according to Claim 38, further comprising:
2 a length of a distance vector determined from the term vector and the
3 centroid; and
4 the scorer computing the score for the scored query as a product of a
5 number of stored documents with which the stored query is associated and the
6 distance vector length.

1 40. A system according to Claim 25, further comprising:
2 a selector selecting the stored query associated with the matching at least
3 one stored document relative to a threshold.

1 41. A system according to Claim 25, further comprising:
2 a sorter sorting the set of query refinements.

1 42. A system according to Claim 25, further comprising:
2 a presenter presenting the set of query refinements.

1 43. A system according to Claim 25, further comprising:

2 a set of supplemental query refinements negating each term in the set of
3 query refinements not present in the search query and using the negated terms in
4 combination with the search query as at least one supplemental query refinement.

1 44. A system according to Claim 25, further comprising:
2 an association database maintaining the pairings.

1 45. A system according to Claim 25, wherein at least one of each such
2 stored document and each such search document is specified as at least one of a
3 Uniform Resource Locator (URL), hyperlink, anchor, and document excerpt.

1 46. A method for providing search query refinements, comprising:
2 associating a stored query and a stored document as a logical pairing and
3 assigning a weight to the logical pairing;
4 issuing the search query and producing a set of search documents;
5 matching at least one search document to at least one stored document and
6 retrieving the stored query and the assigned weight associated with the matching
7 at least one stored document;
8 forming at least one cluster based on the stored query and the assigned
9 weight associated with the matching at least one stored document; and
10 scoring the stored query associated with the matching at least one stored
11 document for the at least one cluster relative to at least one other cluster and
12 suggesting at least one such scored search query as a set of query refinements.

1 47. A method according to Claim 46, further comprising:
2 selecting one such search document chosen from among the set of search
3 documents responsive to the search query issuance as the at least one such search
4 document.

1 48. A method according to Claim 46, further comprising:
2 selecting the set of search documents as the at least one such search
3 document.

1 49. A method according to Claim 46, further comprising:

2 tracking previous search queries; and
3 regenerating a set of previous search documents produced by the previous
4 search queries as the at least one such search document.

1 50. A method according to Claim 46, further comprising:
2 associating at least one cached document and one or more cached queries
3 as a cached pairing; and
4 inverting each cached pairing to associate at least one cached query and
5 one or more cached documents as the at least one such search document.

1 51. A method according to Claim 46, further comprising:
2 for each such pairing, estimating relevancy to the stored query for the
3 stored document as the weight assigned to the pairing.

1 52. A method according to Claim 51, further comprising:
2 summing each such assigned weight for a plurality of pairings
3 corresponding to the stored query and the stored document.

1 53. A method according to Claim 46, wherein each stored query
2 comprises one or more terms, further comprising:
3 computing a term vector comprising the terms in the stored query
4 associated with the matching at least one stored document;
5 determining a distance determined for the term vector; and
6 forming the at least one cluster relative to the distance.

1 54. A method according to Claim 53, further comprising:
2 normalizing the term vector.

1 55. A method according to Claim 53, further comprising:
2 computing a length of the term vector in multi-dimensional space with
3 each dimension equaling a sum of the weights of the term in a set of associated
4 stored queries.

1 56. A method according to Claim 53, further comprising:

2 assigning a relevance score to the at least one search document; and
3 ranking the at least one cluster relative to the at least one other cluster by
4 the relevance score associated with the matching at least one search document and
5 a number of the matching at least one search document.

1 57. A method according to Claim 56, further comprising:
2 selecting one of more of the ranked at least one cluster as potential
3 refinement clusters based on the rankings.

1 58. A method according to Claim 57, further comprising:
2 computing a centroid as a weighted center for each such potential
3 refinement cluster; and
4 scoring the stored query associated with the matching at least one stored
5 document for the potential refinement cluster relative to the centroid.

1 59. A method according to Claim 58, further comprising:
2 computing the centroid as a normalized sum of a product of the term
3 vector for each stored query and the relevance score associated with the matching
4 at least one search document.

1 60. A method according to Claim 59, further comprising:
2 determining a length of a distance vector from the term vector and the
3 centroid; and
4 computing the score for the scored query as a product of a number of
5 stored documents with which the stored query is associated and the distance
6 vector length.

1 61. A method according to Claim 46, further comprising:
2 selecting the stored query associated with the matching at least one stored
3 document relative to a threshold.

1 62. A method according to Claim 46, further comprising:
2 sorting the set of query refinements.

1 63. A method according to Claim 46, further comprising:
2 presenting the set of query refinements.

1 64. A method according to Claim 46, further comprising:
2 negating each term in the set of query refinements not present in the
3 search query and using the negated terms in combination with the search query as
4 at least one supplemental query refinement.

1 65. A method according to Claim 46, further comprising:
2 maintaining the pairings in a database.

1 66. A method according to Claim 46, further comprising:
2 specifying at least one of each such stored document and each such search
3 document as at least one of a Uniform Resource Locator (URL), hyperlink,
4 anchor, and document excerpt.

1 67. A computer-readable storage medium holding code for performing
2 the method according to Claim 46.

1 68. An apparatus for providing search query refinements, comprising:
2 means for associating a stored query and a stored document as a logical
3 pairing and means for assigning a weight to the logical pairing;
4 means for issuing the search query and means for producing a set of
5 search documents;
6 means for matching at least one search document to at least one stored
7 document and means for retrieving the stored query and the assigned weight
8 associated with the matching at least one stored document;
9 means for forming at least one cluster based on the stored query and the
10 assigned weight associated with the matching at least one stored document; and
11 means for scoring the stored query associated with the matching at least
12 one stored document for the at least one cluster relative to at least one other
13 cluster and means for suggesting at least one such scored search query as a set of
14 query refinements.

1 69. A system for integrating query refinement candidates, comprising:
2 a matcher matching at least one search document retrieved responsive to a
3 query to one or more stored documents associated with a stored query and weight
4 and matching at least one further search document retrieved responsive to a
5 candidate query to the one or more stored documents;
6 a cluster forming at least one cluster based on the stored query and weight
7 associated with each stored document matched responsive to the query and
8 forming at least one further cluster based on the stored query and weight
9 associated with each stored document matched responsive to the candidate query;
10 a combiner combining the at least one cluster and the at least one further
11 cluster; and
12 a scorer scoring the stored query for the combined cluster relative to at
13 least one other cluster as a potential query refinement suggestion.

1 70. A system according to Claim 69, further comprising:
2 a set of candidate query refinements comprising at least one such
3 candidate query.

1 71. A system according to Claim 70, further comprising:
2 an evaluator assigning at least one such candidate query to the at least one
3 cluster.

1 72. A system according to Claim 71, further comprising:
2 a builder creating an orthogonal set of candidate query refinements
3 comprising at least one such unassigned query candidate.

1 73. A method for integrating query refinement candidates, comprising:
2 matching at least one search document retrieved responsive to a query to
3 one or more stored documents associated with a stored query and weight;
4 forming at least one cluster based on the stored query and weight
5 associated with each stored document matched responsive to the query;

6 matching at least one further search document retrieved responsive to a
7 candidate query to the one or more stored documents;
8 forming at least one further cluster based on the stored query and weight
9 associated with each stored document matched responsive to the candidate query;
10 combining the at least one cluster and the at least one further cluster; and
11 scoring the stored query for the combined cluster relative to at least one
12 other cluster as a potential query refinement suggestion.

1 74. A method according to Claim 73, further comprising:
2 assembling a set of candidate query refinements comprising at least one
3 such candidate query.

1 75. A method according to Claim 74, further comprising:
2 assigning at least one such candidate query to the at least one cluster.

1 76. A method according to Claim 75, further comprising:
2 creating an orthogonal set of candidate query refinements comprising at
3 least one such unassigned query candidate.

1 77. A computer-readable storage medium holding code for performing
2 the method according to Claim 73.

1 78. An apparatus for integrating query refinement candidates,
2 comprising:
3 means for matching at least one search document retrieved responsive to a
4 query to one or more stored documents associated with a stored query and weight;
5 means for forming at least one cluster based on the stored query and
6 weight associated with each stored document matched responsive to the query;
7 means for matching at least one further search document retrieved
8 responsive to a candidate query to the one or more stored documents;
9 means for forming at least one further cluster based on the stored query
10 and weight associated with each stored document matched responsive to the
11 candidate query;

- 12 means for combining the at least one cluster and the at least one further
13 cluster; and
14 means for scoring the stored query for the combined cluster relative to at
15 least one other cluster as a potential query refinement suggestion.